

DOI: 10.51981/2588-0039.2021.44.031

АВТОМАТИЧЕСКАЯ ИНТЕРПРЕТАЦИЯ ВЫСОКОШИРОТНЫХ ИОНОГРАММ CADI: ОБНАРУЖЕНИЕ СЛОЕВ E, ES, F1 И F2

С.А. Долгачева, А.В. Николаев, Л.Н. Макарова

ФГБУ «Арктический и антарктический научно-исследовательский институт»,
Санкт-Петербург, Россия
E-mails: dolgachyova2010@yandex.ru, demosfen.spb@gmail.com, lumak@aari.nw.ru

Аннотация. Настоящее исследование является продолжением цикла работ, посвященных внедрению искусственных нейронных сетей в автоматическую обработку высокоширотных ионограмм вертикального зондирования ионосферы. Возможности разработанных ранее нейросетевых моделей были ограничены определением типов Es и F2 ионосферных слоев. К настоящему моменту удалось обучить искусственные нейронные сети различать слои E, Es, F1 и F2, типы слоя Es, тип диффузии слоя F2 (F, P, Q, L) и определять их числовые параметры. Эффективность моделей оценивалась с помощью тестовой выборки, содержащей примерно 60000 ионограмм. Точность определения слоев E, Es, F1, F2 с помощью наиболее эффективной модели в настоящий момент достигает 91%.

Введение

На протяжении последних десятилетий ФГБУ «ААНИИ» ведет непрерывные сетевые геофизические наблюдения, в состав которых входит мониторинг состояния высокоширотной ионосферы. Порядка 75% всех ионограмм, поступающих с 15-минутным разрешением в базу данных Полярного Геофизического Центра (ПГЦ) отдела Геофизики, не обрабатываются наблюдателями и хранятся в исходном виде. Поскольку существующие программы автоматической обработки ионограмм имеют ограниченные возможности и не предоставляют удовлетворительный уровень точности [1, 2], нами были разработаны модели на основе нейронных сетей, способные распознать слои Es и F2 и оценить их параметры [3]. В данной работе к слоям Es и F2 предыдущих версий моделей были добавлены слои E и F1, а также дополнительные характеристики слоев Es (тип слоя - f, l, c, h, r, a) и тип диффузии слоя F2 (F, P, Q, L). Числовые параметры, определяемые улучшенными моделями, следующие: (1) критические частоты обыкновенной волны (foE, foEs, foF1, foF2); (2) минимальные действующие высоты (hE, hEs, hF1, hF2); (3) наименьшая наблюдаемая частота отражения от ионосферы (fmin); (4) экранирующая частота слоя Es (fbEs); (5) высота максимума слоя Es (hmEs); (6) наибольшая частота отражения от области F (fx1); (7) предельная частота рассеянной обыкновенной компоненты слоя F (fo1); (8) высота и частота для коэффициента M3000 F1 и F2 слоев (M3000f F1, M3000h F1, M3000f F2 и M3000h F2).

Для обучения (тренировки) нейросетей используются результаты ручной обработки ионограмм. Чтобы достичь максимального разнообразия обучающей базы данных, были отобраны ионограммы со всех станций для четырех сезонов при различном уровне геомагнитной активности. Распознавание ионограмм производится с помощью глубоких нейронных сетей [5, 6], решающих задачи классификации (распознавание слоев, типа слоя Es и типа диффузии слоя F2) и регрессии (числовые параметры, высоты и частоты слоев). В текущем исследовании используется свёрточная нейронная сеть с тремя слоями свёртки и тремя линейными слоями [5]. Более подробная информация о принципах построения моделей и решении задач классификации и регрессии содержится в работе [3].

Обучение моделей и результаты

Обработка ионограммы относится к классической задаче распознавания изображений (задачи компьютерного зрения). Распознавание ионограммы состоит из двух ключевых этапов – выделения признаков из данных и классификации (или регрессии) на их основе. Модели обучались всем классам или всем числовым параметрам сразу, поэтому был собран совокупный обучающий и тестовый набор данных, содержащий примерно одинаковое количество ионограмм всех классов с квазиравномерным распределением по всем определяемым высотам и частотам. Размер тренировочного и тестового наборов составил 283 901 и 60 626 ионограмм соответственно. После обучения и тестирования моделей разных конфигураций были получены три модели с наилучшими показателями точности: (1) глобальная классификационная модель; (2) регрессионная модель для всех параметров и (3) модель классификации типов спорадических слоев Es, с точностью выше, чем у глобальной классификационной модели.

Показатели точности моделей представлены в таблицах 1-3. Для оценки качества классификации на каждом из классов были выбраны следующие метрики: (1) доля правильных ответов «ассигасу» - отношение правильно найденных и не найденных экземпляров класса ко всему объему данных; (2) точность «precision» - отношение правильно найденных экземпляров класса ко всем найденным экземплярам класса; (3) полнота recall - отношение правильно найденных экземпляров класса ко всем имеющимся экземплярам класса. Для оценки качества регрессии выбрана средняя абсолютная ошибка (MAE). В таблице 1 приведены результаты оценки качества распознавания ионограмм глобальной классификационной и регрессионной моделями.

Как было упомянуто выше, обучающий набор данных выбирался таким образом, чтобы значения всех параметров были равномерно распределены. Однако для некоторых типов спорадического слоя (I, h и a), а также типов диффузии слоя F2, набор данных сбалансировать не удалось. Так, например, доля ионограмм с диффузией слоя F2 F-типа составила 9.93%, P типа 0.07%, Q типа 1.27%, L типа 3.91%, при этом 84.82% ионограмм не содержали слоя F2 или диффузии слоя F2. Как видно из таблицы 2, типы диффузии P и Q моделью не обнаруживаются вследствие малого количества ионограмм с таким типом диффузии. В этом случае модель стремится определить отсутствие диффузии слоя F2.

Для несбалансированных наборов данных при обучении модели используются метрики, не зависящие от соотношения классов, например, F-мера, среднее гармоническое precision и recall (максимальное значение 1). Однако использование F-меры при обучении глобальной классификационной модели не помогло добиться удовлетворяющей точности определения типа диффузии слоя F2 и типа слоя Es, поэтому для решения задачи распознавания типа слоя Es была разработана модель с отдельным набором данных, сбалансированным по всем классам. Набор данных для обучения этой модели содержит только ионограммы с наличием Es, около 5000 ионограмм каждого определяемого типа. Тестовый набор данных содержит порядка 3000 ионограмм каждого типа. Результаты оценки качества модели классификации слоя Es приведены в таблице 3, из которой видно, что высокая точность определения чужих классов улучшает показатель ассигасу, но при этом модель стремится все классы определить как чужие. На рисунке 1 приведена матрица распознанных классов тестового набора данных. Что касается слоя F2, на данный момент модель для определения типа его диффузии с удовлетворяющей точностью получить не удалось.

Пример определения слоя и его критической частоты представлен на рисунке 2.

Таблица 1. Точность определения слоев ионосферы и их основных параметров.

Тип слоя	Accuracy %	Precision %	Recall %	F-мера	MAE fo, МГц	MAEh, км	MAE M3000 f, МГц	MAE M3000 h, км
E	88	78	67	0.72	0.39	6.2	–	–
Es	85	90	88	0.89	0.66	8.25	–	–
F1	91	81	65	0.72	0.43	25	0.5	27
F2	89	92	88	0.90	0.4	31	0.4	32

Таблица 2. Точность определения дополнительных параметров ионосферы и типов диффузии слоя F2.

Параметр	MAE	Тип диффузии F2	F-мера
f min, MHz	0.32	отсутствие диффузии	0.95
fb Es, MHz	0.47	F	0.34
foI, MHz	0.27	P	-
fxI, MHz	0.45	Q	-
hm, km	8.5	L	0.05

Таблица 3. Точность определения типа слоя Es.

Метрика \ Тип слоя Es	f	l	C	h	r	a
Accuracy, %	81	77	82	82	81	81
Precision, %	38	38	46	44	43	47
Recall, %	30	48	46	44	43	47
F-мера	0.33	0.43	0.46	0.44	0.43	0.47

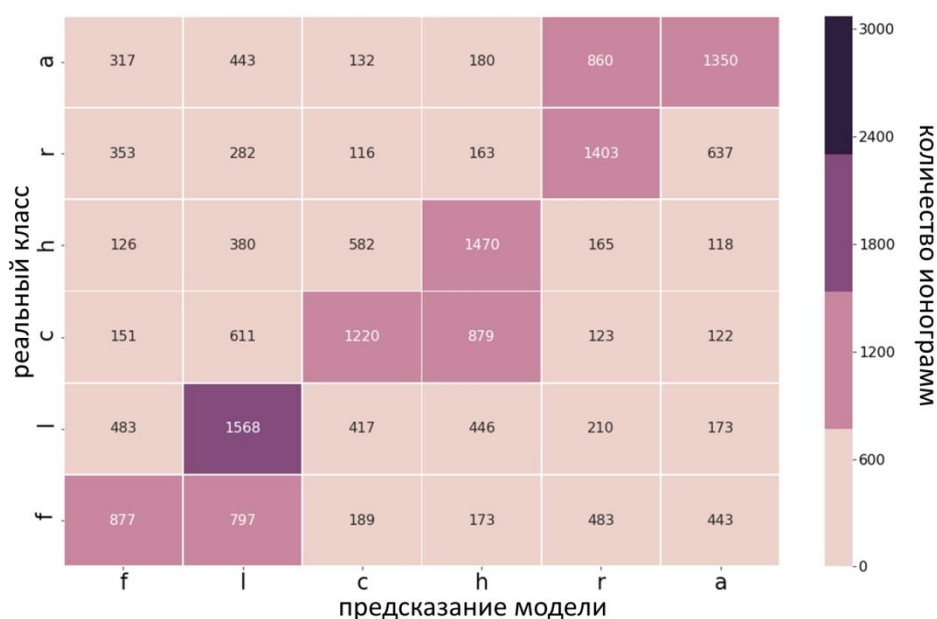


Рисунок 1. Матрица распознанных типов слоя Es на тестовом наборе данных.

Заключение

Автоматическая интерпретация высокоширотных ионограмм является сложной задачей, в отличие от задачи обработки среднеширотных ионограмм. Внедрение искусственных нейронных сетей имеет большой потенциал в вопросе улучшения качества распознавания характерных структур на ионограммах и получения их числовых характеристик. В данной работе приведены результаты обработки 60 626 ионограмм нейросетевыми моделями с расширенными возможностями. Доля правильных ответов (ассигасу) наиболее эффективной модели при определении слоёв E, Es, F1, F2 составляет от 85% до 91%. Метрика качества (F-мера) определения типа слоя Es составляет от 0.33 до 0.47, лучший результат определения типа диффузии F2 слоя – 0.34. В настоящий момент средняя абсолютная ошибка определения разного типа частот составляет от 0.27 до 0.66 МГц; средняя ошибка определения высот E области от 6.2 до 8.5 км, F области от 25 до 32 км.

Благодарности

Работа выполнена в рамках научно-исследовательских работ по теме 6.1 «Развитие и модернизация технологий мониторинга геофизической обстановки над территорией Российской Федерации и Арктики».

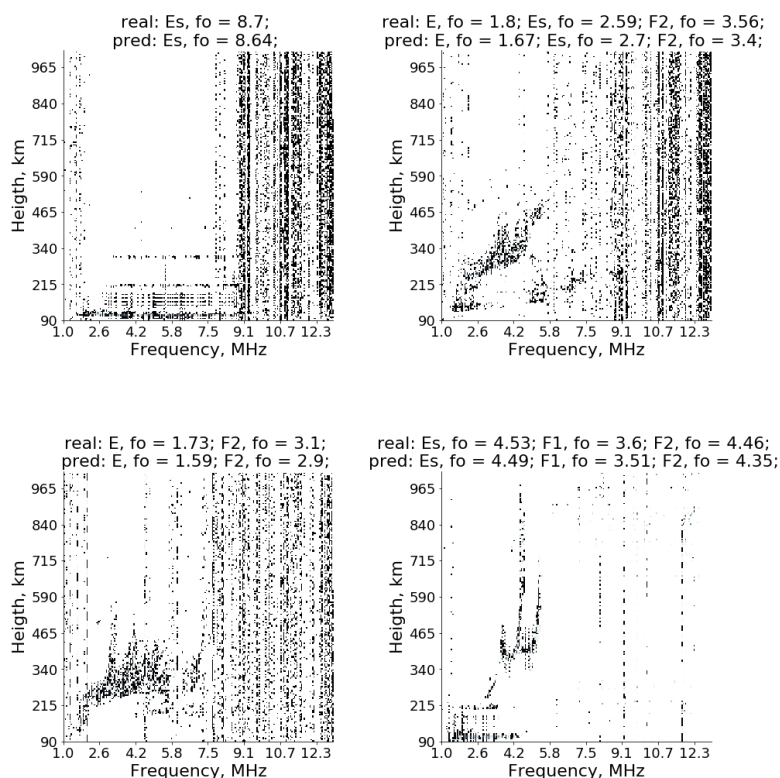


Рисунок 2. Пример распознавания реальной ионограммы: результаты определения слоя и его критической частоты. Real – результаты обработки ионограммы вручную, pred – результат моделей.

Литература

1. Выставной В.М., Макарова Л.Н., Широчков А.В., Егорова Л.В. Исследования высокоширотной ионосферы методом вертикального зондирования с использованием современного цифрового ионозонда cad1. Гелиогеофизические исследования, № 4, С.1–10, 2013.
2. Выставной В.М., Егорова Л.В., Макарова Л.Н., Широчков А.В. Измерения параметров высокоширотной ионосферы методом вертикального зондирования с помощью современного цифрового ионозонда. Мир Измерений, № 10, С. 49–55, 2014.
3. Долгачева С.А., Макарова Л.Н., Николаев А.В. Обработка ионограмм высокоширотных станций вертикального зондирования с использованием нейронных сетей: Es и F2 слои. Physics of Auroral Phenomena, Т. 43, С. 105–108, 2020.
4. Руководство URSI по интерпретации и обработке ионограмм. Пер. с англ., Междунар. науч. радиосоюз, Сов. нац. ком., М.: Наука, 1978, 342 с.
5. Николенко С., Кадури А., Архангельская Е. Глубокое обучение. СПб.: Питер, 2018, 480 с.
6. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Deep Residual Learning for Image Recognition. arXiv:1512.03385, 2015.
7. Pan S.J., Yang Q. A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 22(10), 1345–1359, 2010.